Sample text data

docs <- Corpus(VectorSource(c("The article aims to develop interpretable Machine Learning models using R statistical programming language for malaria risk prediction in Kenya, emphasizing leveraging Explainable AI (XAI) techniques to support targeted interventions and improve early detection mechanisms. The methodology involved using synthetic data with 1000 observations, employing over-sampling to address class imbalance, utilizing two machine learning algorithms (Random Forest and Extreme Gradient Boosting), applying cross-validation techniques, Hyper-parameter tuning and implementing feature importance and SHAP (Shapley Additive Explanations) for model interpretability. The findings revealed that Random Forest outperformed Extreme Gradient Boosting with 98% accuracy. Critical prediction features included clinical symptoms such as nausea, muscle aches, and fever, plasmodium species identification, and environmental factors like rainfall and temperature. Both models demonstrated strong sensitivity in detecting malaria cases. This promotes trust in model predictions by clearly outlining the decision process for individual outcomes. The research concluded that integrating Explainable AI into malaria risk prediction represents a transformative approach to public health management. Through providing transparent, interpretable models, the research offers a robust, data-driven approach to predicting malaria risks, potentially empowering healthcare providers and policymakers to deploy resources more effectively and reduce the disease burden in endemic regions.")))